Lecture 12
Simulate Random Variables I
Instructor: Hongxiao Zhu
Quiz Questions:

• Which of the following are true?

  (a) one can generate true random numbers using a computer.

  (b) there is no way to reproduce a sequence of random numbers generated by a computer program.

  (c) one can simulate from all distributions by using Unif(0, 1) random numbers.
• In previous lecture, we learned congruential generators for Unif(0,1), and used Unif(0,1) random numbers to calculate the area of an irregular shape.
Exercise from last lecture

Following Example 2, calculate the area of a half-disk defined by:

\[ y = \sqrt{1 - x^2}, \quad x \in [-1, 1]. \]

(1) Increasing the total number of monte carlo samples from \( N=500 \) to \( N=5000 \). Compare the results.

(2) Plot the Monte Carlo samples on top of the disk area one by one, and color those points inside and outside the half-disk differently.
Today we will learn:

• Given Unif (0, 1) random numbers, how to simulate random numbers that follow certain distribution?
Discrete Random Variables
Discrete Random Variables

Probability Mass Function

Cumulative Distribution Function
Generate random samples from discrete distributions

• X takes values in \{1, 3, 7\}, with pmf

\[
\begin{align*}
Pr(x=1) &= 0.2 \\
Pr(x=3) &= 0.5 \\
Pr(x=7) &= 0.3
\end{align*}
\]

• Assume we only know how to simulate runif(0,1). How do we generate samples for X?
If the pmf is known...

All masses constitute a segment with length 1
If the pmf is known...

The Segmentation Approach:

1. Simulate $u \sim \text{Unif}(0, 1)$

2. Set $X = \begin{cases} 
1, & \text{if } u < 0.2. \\
3, & \text{if } 0.2 \leq u < 0.7 \\
7, & \text{if } 0.7 \leq u < 1. 
\end{cases}$
If the CDF function $F(x)$ is known.

All jumps constitute a CDF with height 1.
If the CDF function $F(x)$ is known.

The CDF Approach:

```r
# given $U \sim U(0,1)$
X <- 0
while (F(X) < U) {
  X <- X + 1
}
```
R functions

To simulate binomial, geometric, negative-binomial or Poisson rv’s in R, use `rbinom`, `rgeom`, `rnbinom` or `rpois`. For simulating other discrete rv’s R provides `sample(x, size, replace = FALSE, prob = NULL)`

The inputs are

- `x` A vector giving the possible values the rv can take;
- `size` How many rv’s to simulate;
- `replace` Set this to `TRUE` to generate an iid sample, otherwise the rv’s will be conditioned to be different from each other;
- `prob` A vector giving the probabilities of the values in `x`. If omitted then the values in `x` are assumed to be equally likely.
Examples:

**Example 1.** Generate 100 samples for random variable X with the following pmf. Use the segmentation approach. Compare the histogram with that produced using sample() function.

<table>
<thead>
<tr>
<th>x</th>
<th>1</th>
<th>0</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pr(X=x)</td>
<td>0.6</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Example 2.** Redo example 1 using the CDF approach. Use the while statement. Compare the histogram with the sample() output.
Continuous Random Variables
Continuous Random Variables

**Probability Density Function**

\[ p(x) \]

\[ \int_a^b p(x) \, dx \]

**Cumulative Distribution Function**

\[ F_X(x) \]
If we know the CDF...

• Let $U \sim \text{Unif}(0, 1)$, then $u$ take values between $(0, 1)$.
The Inverse CDF approach.

- Let $U \sim \text{Unif}(0, 1)$. For any continuous CDF $F(x)$, the random variable $X$ defined by $X = F^{-1}(U)$ has distribution $F$.

**Proof:**
Define $X = F^{-1}(U)$ for $U \sim \text{Unif}(0, 1)$.
We want to show that $X$ follows a distribution with CDF function $F(x)$.

\[
Pr(X \leq x) = Pr(F^{-1}(U) \leq x) = Pr(U \leq F(x)) = F(x)
\]
**Example 3:** Generate $\text{Exp}(\lambda)$ using the inverse CDF method. Let $\lambda = 1$.

- **Step 1:** Find the CDF for $\exp(\lambda)$.

  If $X \sim \exp(\lambda)$, then

  \[
  F_X(x) = 1 - e^{-\lambda x}
  \]

  for $x \geq 0$.

- **Step 2:** Let $u = F_X(x)$, and solve $x$ as a function of $u$.

  \[
  x = F_X^{-1}(u) = -\frac{1}{\lambda} \log (1 - u).
  \]

- **Step 3:** Write code to simulate $X$ using inverse CDF method.
Take home message:

• Three approaches for simulating discrete random variables:
  • The segmentation approach
  • The CDF approach
  • sample()

• One method for simulating continuous random variables
  • Inverse CDF approach
Exercise:

1. Using the segmentation approach to generate 100 samples for random variable \( X \) with the following pmf

\[
\begin{array}{c|ccc}
 x & 2 & 1 & 5 \\
 p_X(x) & 0.3 & 0.1 & 0.6 \\
\end{array}
\]

and compare histogram plot with that produced using samples generated by \texttt{sample()} function.

2. Using the CDF approach to generate 1000 samples from \texttt{Binom(10, 0.5)}. Use \texttt{while} loop. Note that you can use \texttt{pbinom()} as the CDF function. Compare histogram with that produced by sampling from \texttt{rbinom()}.

3. Use the inverse CDF approach to simulate 1000 samples from \texttt{Unif(0,1)}, and compare histogram with that produced by \texttt{runif}().